remote sensing

Communication

Texture-Mapping Error Removal Based on the BRIEF Operator in Image-Based Three-Dimensional Reconstruction

Junxing Yang [1,†], Lu Lu [2,†], Ge Peng [1], He Huang [1], Jian Wang [1] and Fei Deng [3,*]

[1] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China
[2] Zhengzhou Bureau of Natural Resources and Planning, Zhengzhou 450006, China
[3] School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China
[*] Correspondence: fdeng@sgg.whu.edu.cn
[†] These authors contributed equally to this work.

Abstract: In image-based three-dimensional (3D) reconstruction, texture-mapping techniques can give the model realistic textures. When the geometric surface in some regions is not reconstructed, such as for moving cars, powerlines, and telegraph poles, the textures in the corresponding image are textured to other regions, resulting in errors. To solve this problem, this letter proposes an image consistency detection method based on the Binary Robust Independent Elementary Features (BRIEF) descriptor. The method is composed of two parts. First, each triangle in the mesh and its neighboring triangles are sampled uniformly to obtain sampling points. Then, these sampled points are projected into the visible image of the triangle, and the corresponding sampled points and their RGB color values are obtained on the corresponding image. Based on the sampled points on these images, a BRIEF descriptor is calculated for each image corresponding to that triangle. In the second step, the Hamming distance between these BRIEF descriptors is calculated, outliers are removed according to the method, and noisy images are also removed. In addition, we propose adding semantic information to Markov energy optimization to reduce errors further. The two methods effectively reduced errors in texture mapping caused by objects not reconstructed, improving the texture quality of 3D models.

Keywords: 3D reconstruction; texture mapping; BRIEF descriptors; outlier detection

check for updates

Citation: Yang, J.; Lu, L.; Peng, G.; Huang, H.; Wang, J.; Deng, F. Texture-Mapping Error Removal Based on the BRIEF Operator in Image-Based Three-Dimensional Reconstruction. *Remote Sens.* **2023**, *15*, 536. https://doi.org/10.3390/rs15020536

Academic Editors: San Jiang, Xiongwu Xiao and Wanshou Jiang

Received: 29 November 2022
Revised: 5 January 2023
Accepted: 6 January 2023
Published: 16 January 2023

1. Introduction

Image-based three-dimensional (3D) reconstruction has achieved great success in recent decades. This technology can recover not only the geometric structure of the object surface but also the color information of the object surface from multi-view images. The pipeline for geometric reconstruction by this technology is the structure from motion [1–3], dense matching [4–6], and surface construction [7–9]. However, the images are taken at different times, and there are problems involving weak texture, repeated texture, and moving targets. Furthermore, objects such as power lines, pedestrians, and moving vehicles cannot be reconstructed by multi-view stereo algorithms, and these objects are often textured to other regions during the texture-mapping process [10]. These problems result in errors, and solving them is important for improving the quality of realistic 3D models.

In the field of image-based 3D reconstruction, traditional texture-mapping methods fall into two main categories: the image-fusion-based method [11–13] and the view-selection-based approach [10,14–16]. The former uses image-weighted fusion to reconstruct textures for the mesh. If there is a moving vehicle on the road, some of the visible images of the mesh triangle have captured the moving vehicle, while others have not. If the image-fusion-based method is used for reconstructing texture, fuzzy problems will inevitably occur in the vehicle area. The latter selects an optimal image for each triangle in the mesh by view selection and then maps the corresponding region of that image to the corresponding triangle to obtain the texture. If this type of method includes an image that has not been reconstructed for a triangle of the mesh, an error will occur when the image is mapped to the mesh. Therefore, it is necessary to remove the images containing unreconstructed objects in advance, but the existing methods cannot meet the needs of texture mapping, which is described in detail in the next paragraph. Iteration-based methods [17,18] can be used to reconstruct clear textures, but they are not suitable for large-scale texture mapping because they require too many iterations and are time-consuming. With the development of deep learning technology, we can use neural networks to represent the entire scene and then give a new perspective to render a new image by volume rendering. This has become a salient topic in recent years, especially in the neural radiance field (NeRF) [19–21]. Moreover, this method can be applied to texture mapping to provide real textures for the entire scene. For example, Martin [22] proposed NeRF-W, which can deal with the color consistency of images and remove mobile people. Other researchers [23–25] reconstructed the normal vector, albedo, and BRDF of the object surface so that the entire scene can be relighting. However, how to combine NeRF with traditional texture-mapping technology to generate high-quality textures still needs further exploration, so it is a method worth studying.

There are two main solutions that may be able to account for the influence of non-reconstructed objects on texture-mapping results. One is the method based on image consistency detection. Because only a few images in the visible image list of a single mesh triangle contain objects not reconstructed, consistency detection can be performed in the visible image list based on the color information within the projection of the triangle to eliminate the images containing objects not reconstructed. Sinha et al. [26] and Grammatikopoulos et al. [27] used the median or average value of color information to detect image consistency. Waechter et al. [10] used a mean-shift-based image consistency detection method to remove the images containing objects that have not been reconstructed. These methods can achieve good results, but for the texture mapping of urban 3D models with many objects not reconstructed, these methods still produce many texture-mapping errors. Another approach based on masked images introduces semantic information in the texture-mapping process. This is an effective method that uses deep-learning-based object detection methods [28,29] to detect objects that have not been reconstructed, such as mobile vehicles and pedestrians, and then remove them [30–32]. These deep-learning-based algorithms can only process specific objects based on specific training datasets, such as cars and pedestrians. However, too many objects that have not been reconstructed cannot be detected by deep learning, which also limits the application scope of such methods.

In this article, we propose a BRIEF-based texture-mapping image consistency detection method that assigns a modified BRIEF descriptor [33] to each image to perform outlier detection. This method can eliminate "outlier" images in the visible image of a single mesh triangle list by calculating the Hamming distance of the corresponding descriptors, which is equivalent to eliminating images containing non-reconstructed objects. In addition, we also take the mobile vehicle as an example to show how to introduce semantic information into view selection to reduce texture-mapping errors further.

2. Methods

2.1. Principle of the BRIEF Descriptor

The BRIEF descriptor is a binary descriptor proposed by Calonder et al. [33] to describe image feature points. The BRIEF descriptor method is as follows:

(a) A neighborhood space p of size $S * S$ centered on the feature points is selected. Gaussian kernel convolution smoothing is performed on the neighborhood space p to diminish the effect of noise.

(b) A total of n sets of point pairs (x, y) are selected in the neighborhood space p, where x obeys the Gaussian$(0, 0.04 * S^2)$distribution and y obeys the Gaussian$(x_i, 0.01 * S^2)$ distribution for sampling.

(c) BRIEF descriptors are constructed as follows. First,define τ:

$$\tau(p; x, y) = \begin{cases} 1, \text{ if } p(x) < p(y) \\ 0, \text{ if } p(x) \geq p(y) \end{cases} \tag{1}$$

where $p(x)$ is the pixel intensity value of p at x. If $p(x) < p(y)$, τ is 1; otherwise, it is 0. Then,define the BRIEF descriptor as a binary string of n_d bits containing only values of 1 and 0:

$$f_{n_d}(p) = \sum_{1 \leq i \leq n_d} 2^{i-1} \tau(p; x_i, y_i) \tag{2}$$

After obtaining the BRIEF descriptors according to the (a) and (b) above, we can perform feature matching using the Hamming distance. The authors of BRIEF found through experiments that the effect of 512 was the best, the effect of 128 was less good, and the effect of 256 was slightly worse than that of 512, but with fewer bits. Hamming distance is a concept that represents the number of different characters in corresponding positions of two (same length) strings. This distance is the number of different characters between BRIEF descriptors at the corresponding positions of the feature vector. When performing feature matching, the following principles are followed:

(d) If the Hamming distance between BRIEF descriptors of two feature points is greater than 128, then they must not match.

(e) The pair of feature points with the smallest Hamming distance between the BRIEF descriptors will be matched.

2.2. Proposed BRIEF-Based Method

For objects not reconstructed, we assume that for a particular mesh triangle, only a small portion of its visible image list contains these objects. Most of the images have similar texture information corresponding to that mesh triangle. Therefore, the problem of removing the visibility impact of objects not reconstructed can be transformed into the problem of removing "outlier" images from the list of visible images of the mesh triangle. In this section, we propose a BRIEF-based texture-mapping image consistency detection method, as shown in Figure 1. This method assigns a modified BRIEF descriptor to each image based on the image color values in the mesh triangle projection range. The "outlier" images are removed by calculating the Hamming distance of the corresponding descriptors. This means that images containing non-reconstructed objects, such as moving vehicles, are excluded. The process is as follows:

(a) For each image in the list of visible images of one mesh triangle F_i, the color values of several points within the projection range on the image of mesh triangle F_i and the three mesh triangles that have common borders with it are sampled uniformly. In practice, the extent of the field and the number of sampled points can be determined according to the scale. In this article, seven points are sampled within each mesh triangle, the vertices of the mesh triangle, and the six points sampled on the borders, as shown in Figure 1.

(b) The sampling points obtained in the first step are combined into pairs of points $n_d(x, y)$, and we define τ, which is the same as in Equation (1).
The n_d results are formed into a string $f_{n_d}(F_i)$ from the lowest bit to the highest bit, with the following formula:

$$f_{n_d}(F_i) = \sum_{1 \leq j \leq n_d} 2^{j-1} \tau(F_i; x_j, y_j). \tag{3}$$

Note that the order of the point pairs in each image in the list of visible images for mesh triangle F_i should remain consistent.

(c) The Hamming distance of any two images' feature vectors is calculated. The Hamming distance of two feature vectors is the number of different characters in the corresponding positions.

(d) Outliers are detected using the following method. Define the distance factor r and the scale factor f, and calculate the number of images whose Hamming distance from each image's corresponding feature vector to the corresponding string of other visible images is greater than r. If the number of images is greater than $f * m$ (m is the number of all visible images of the triangle), the current image is removed. In this method, r is taken as half of the maximum Hamming distance between the image corresponding feature vector, and f is taken as 0.6 based on our experimental experience. To ensure the quality of texture mapping, we stopped the detection if the number of images was less than four during the image consistency detection.

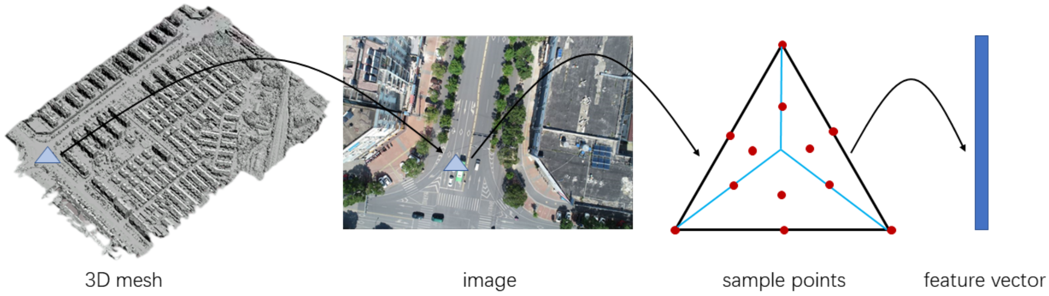

Figure 1. Schematic diagram of feature vector generation. From left to right is a triangle T_i of 3Dmesh, T_i in the projection area R_k of the visible image I_k, the sampling points in the R_k area, and the final feature vector.

2.3. View Selection with Semantic Information

Although the above proposed BRIEF-based method can remove most of the errors in texture mapping, it will fail when encountering some weak texture regions because the brief descriptor depends on the color information of the image. In this section, we take the mobile vehicle as an example to show how semantic information can be introduced into view-selection-based [10] texture-mapping methods to reduce errors. Images containing semantic information can be obtained easily by deep learning methods. The image semantic segmentation technique, however, is not the focus of this letter. Instead, we directly adopted the manual segmentation to obtain a set of vehicle-masked images of the multiview image dataset to show our view selection method with semantic information. The semantic information mask of the image is shown in Figure 2.

The energy function of the Markov random field (MRF) has a data term and a smoothing term. We used the data term to describe the quality of the image selected by a triangle. We introduced the view selection method with vehicle semantic information and defined a new data term. The data term is set to a minimal value if vehicles are in the projection range of the vehicle mask image; otherwise, it takes the gradient amplitude of the triangle surface in the projection range of the image as mvs-texturing. Note that because mvs-texturing only uses gradient data items, the algorithm tends to select images with clear textures, that is, images with large gradients. However, in areas with vehicles, power lines, and other objects, the gradient is often large. Therefore, this kind of data item tends to map these unreconstructed objects to the road surface, thus, causing errors. This is also the problem to be solved in this article. The proposed data term is shown in the following equation:

$$E_{\text{data}}\left(F_i, l_i, m_i\right) = \left\{ \begin{array}{c} \min, F_i \in \text{mask} \\ -\int_{\phi(F_i, l_i)} \left\| \nabla\left(I_{l_i}(p)\right) \right\|^2 dp, F_i \notin \text{mask} \end{array} \right. \quad (4)$$

where m_i is the sum of all pixel values of the mesh triangle F_i in the projection range of the vehicle mask image corresponding to label l_i. When $\text{m}_i > 0$, there are vehicles in the projection range of the mesh triangle in the vehicle mask image; when $m_i = 0$, there are no vehicles. In addition, the method in this article uses the Potts model [10] as the smoothing term as shown in Equation (5).

$$E_{\text{smooth}}\left(F_i, F_j, l_i, l_j\right) = \begin{cases} 0, & l_i = l_j \\ 1, & l_i \neq l_j \end{cases} \tag{5}$$

where F_i and F_j represent the two adjacent triangles, and l_i and l_j represent the image tags selected by F_i and F_j.

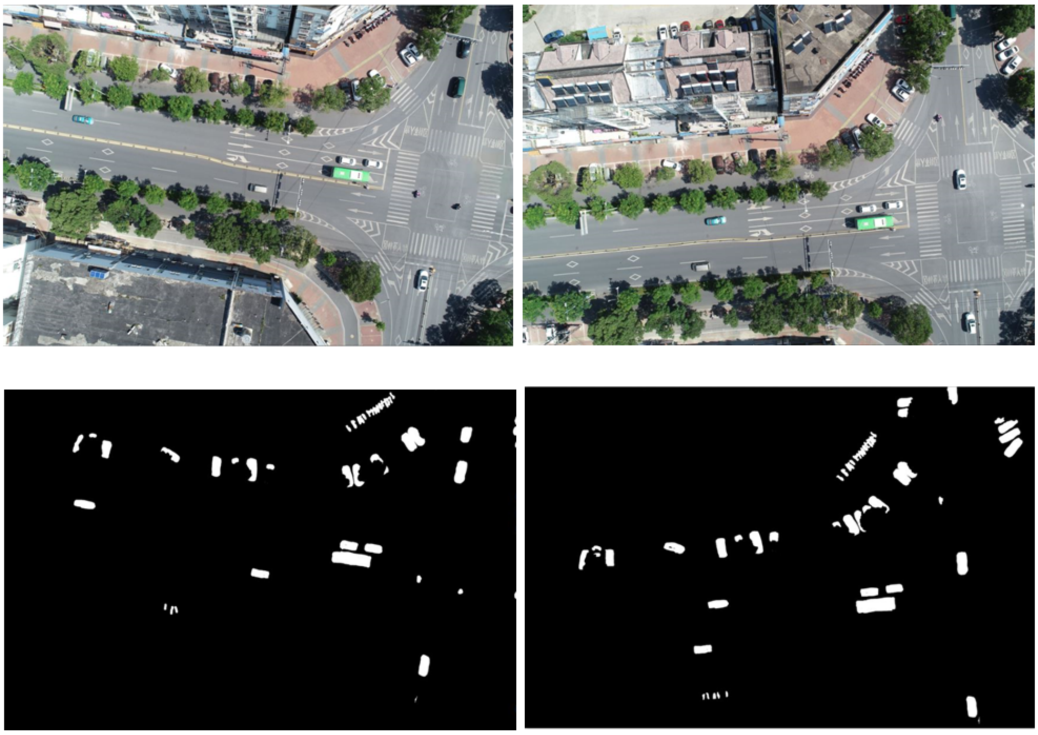

Figure 2. Schematic diagram of vehicle mask images. The first row is the original image; the second row is the moving vehicle mask of the original image.

A multi-labeled graph is constructed as shown in Figure 3. The blue area of the multi-label graph corresponds to the undirected graph generated by the 3D model, and the nodes of the undirected graph correspond to the triangular surface of the 3D model. There are also k terminal nodes in the figure, and each terminal node corresponds to an image tag. In addition, a multi-label graph contains two kinds of edges: the edges of an undirected graph are called n-links, which represent the adjacent relationship between two triangular faces, and the edge connecting the terminal node and the vertex of the undirected graph is called a t-link, which means that the triangular surface is visible on the corresponding image. Each edge in the multi-label graph has weight, and the weight of these edges is determined by the MRF energy function: the data item of the MRF energy function determines the weight of the t-link edge in the multi-label graph, and the smooth item determines the weight of the n-link edge in the multi-label graph, as described in Section 2.3. When we take these edges as water pipes, the process of obtaining optimal image labels for each triangle is equivalent to finding the maximum water flow between nodes. According to the maximum flow minimum cut theorem, when n-link edges are located in the maximum flow and the terminal nodes are divided, the cut of these edges is the minimum cut. In this way, a bridge between graph cut theory and view selection is built through the MRF energy function.

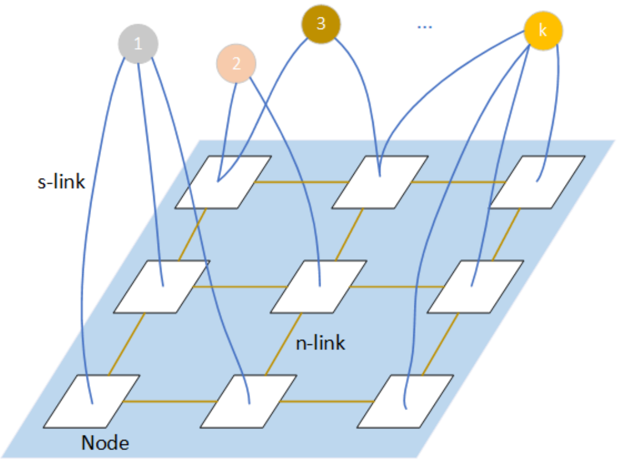

Figure 3. Schematic diagram of multi-label graph cut.

We solved the texture-mapping view selection with semantic information by minimizing the Markov random field (MRF) energy function. The view selection method introducing semantic information has two main parts: (1) constructing the MRF energy function with semantic information and (2) graph cut solving [34]. The swapping algorithm and α-expansion algorithm are only different in the way of label adjustment, which is a key step of the two algorithms. The processes of the two algorithms can be summarized as follows: first, generate initial tags randomly. Then, each expansion (swap) adjusts one (two) labels to reduce the value of the energy function, traversing all possible label combinations. If, under all combinations, the value of the energy function can no longer be reduced by readjusting the label, the locally optimal solution is obtained.

3. Experiments

3.1. Datasets

We compared and analyzed the texture-mapping results without image consistency detection, using BRIEF-based image consistency detection and mvs-texturing [10] on three datasets, as shown in Figure 4 and Table 1. Except for the image consistency detection method, all other parameter settings remained the same.

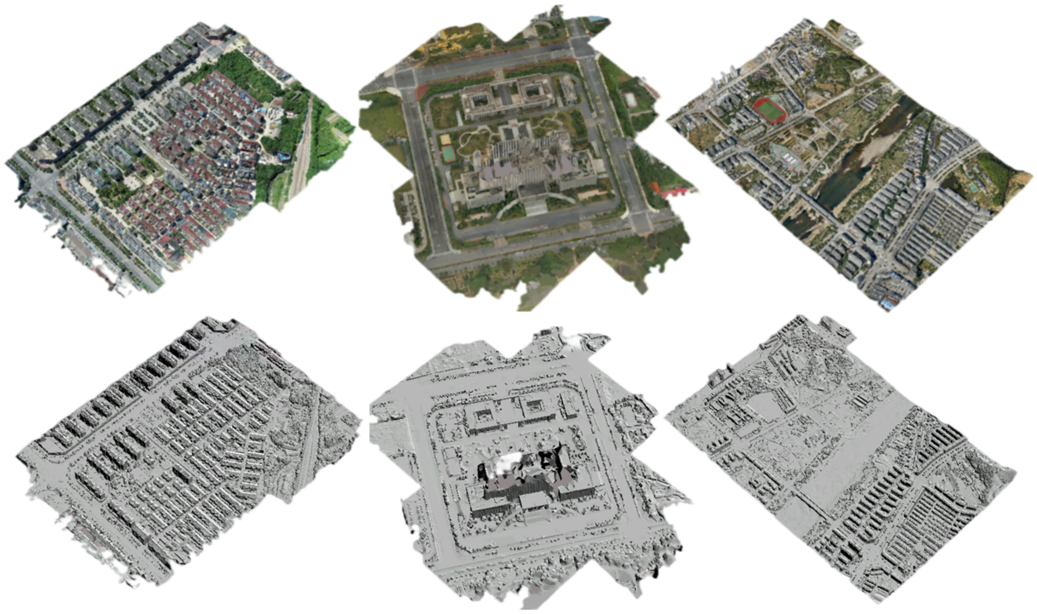

Figure 4. Three datasets used in experiments. The datasets from left to right are TJH-0078, TJH-008812, and TJH-009634.

Table 1. Details of the three datasets.

Datasets	Platform	No. of Images	Size	Altitude (m)	Focal (mm)
TJH-0078	Fixed wing UAV	323	5472 ∗ 3648	135	8.440
TJH-008812	Fixed wing UAV	511	6000 ∗ 4000	118	25.766
TJH-009634	Fixed wing UAV	687	5566 ∗ 3713	330	50.000

3.2. Results and Discussion

Figure 5 shows a comparison of the texture-mapping results, and the errors on the road area are marked with red boxes. The texture-mapping results without image consistency detection had many moving vehicle textures on the road area, and the texture-mapping results with mvs-texturing reduced the number of moving vehicle textures. However, there are still some moving vehicles. In contrast, the results of this letter using the BRIEF-based image consistency method had few moving vehicle textures, which proves that the proposed BRIEF-based image consistency method is more effective than mvs-texturing. Mvs-texturing considered only the average color values in the triangular projection range. In contrast, the BRIEF-based image consistency method used multiple point pairs in the triangular projection range to form a binary string, which better reflected the texture information. In addition, note the BRIEF-based image consistency method indiscriminately rejected “outlier” images; the effectively removed not only erroneous textures of moving vehicles but also erroneous textures of other moving objects. Note, for example, the pedestrian shown in Figure 5.

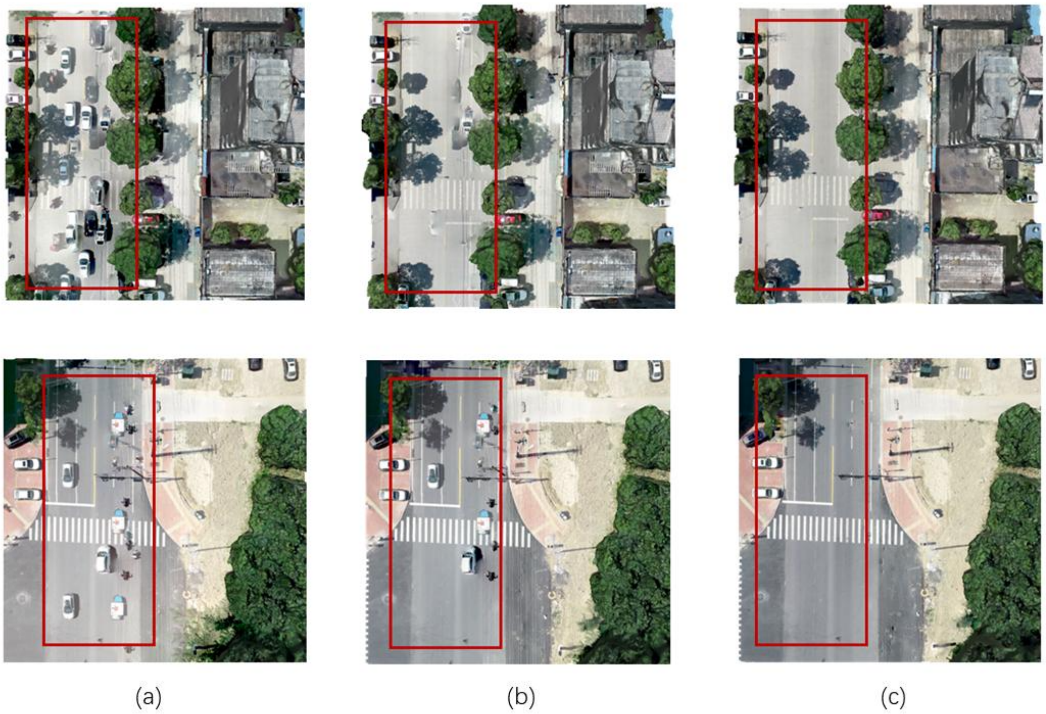

Figure 5. Comparison of the results of the three methods on dataset TJH-0078: (**a**) results of experiments without image consistency detection; (**b**) results of the method of mvs texturing; and (**c**) results of our method.

Similarly, this BRIEF-based image consistency detection method essentially eliminated “outlier” images from the list of visible images in the mesh triangle. For small objects, such as power lines and streetlights that are not geometrically reconstructed, only the images from a specific viewpoint in the list of visible images in a specific triangle contained such small objects. Most of the images from a specific viewpoint did not contain these objects, so the images containing them also could be identified as “outlier” and be removed. Figure 6 compares the texture-mapping results of road wires and shows the zoomed-in portion of the road containing wires. Compared with the results obtained by mvs-texturing, the texture-mapping results of BRIEF-based image consistency detection effectively reduced the errors caused by wires. This shows that our method has better generalization than some deep learning method [30–32] that can only remove vehicles.

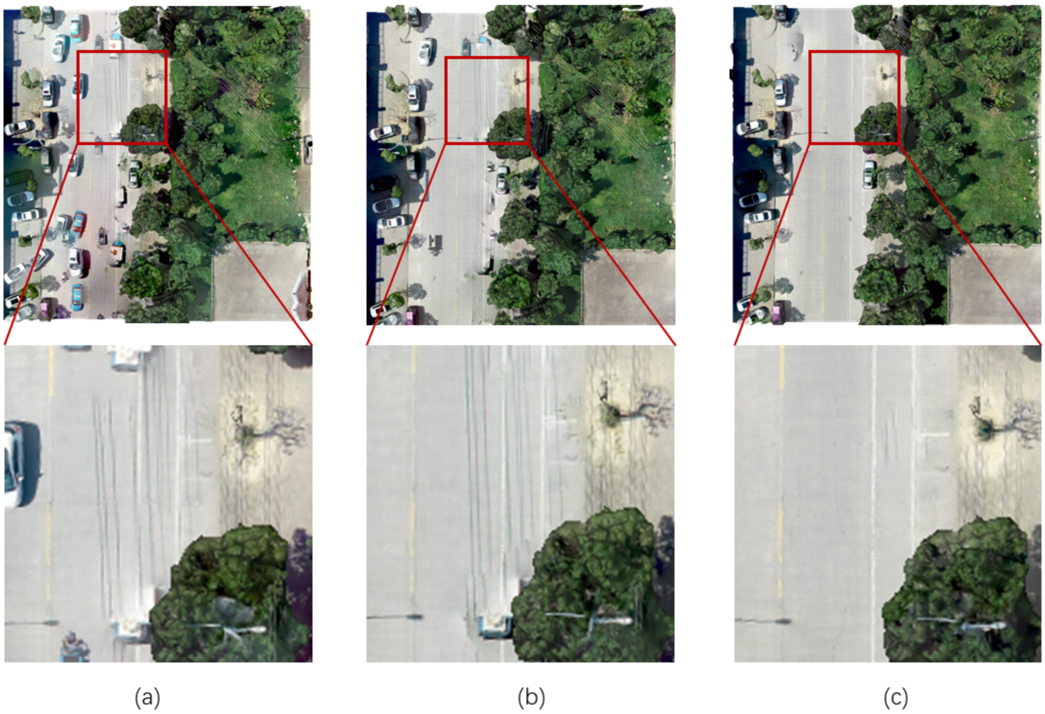

Figure 6. Comparison of texture-mapping results of the road wires of in dataset TJH-0078. (**a**) results of experiments without image consistency detection; (**b**) results of the method of mvs-texturing; and (**c**) results of our method.

We also performed a quantitative comparison of the methods. However, because there is no unified quantitative evaluation criteria for removing texture-mapping errors, this letter counts the number of errors in texture-mapping as a quantitative evaluation criterion. The fewer errors, the better the method. The numbers of mvs-texturing errors on the three datasets are 65, 41, and 205, respectively, and the numbers of our method's errors are 13, 11, and 67, respectively, as shown in Table 2. It can be seen that the number of errors of the method proposed method in this article is far fewer than those of mvs-texturing, which also indicates that the proposed image consistency detection method based on BRIEF used in this article can more effectively avoid errors and, thus, improve the semantic accuracy of texture-mapping results.

Table 2. Statistics of texture-mapping errors of mvs-texturing and ours on three datasets. The fewer the errors, the better the algorithm.

	TJH-0078	**TJH-008812**	**TJH-009634**
mvs-texturing	65	41	205
ours	13	11	67

In addition, because the occlusion detection model does not include the 3D models of adjacent tiles, the occlusion detection in the texture-mapping process may fail, which will cause the triangle surface corresponding to the background object to include the wrong texture of the foreground object within the projection range of some images in its visible image list. Like the small volume objects that have not been reconstructed, only the images with specific viewing angles in the triangle visible image list contain foreground objects, so the images containing foreground objects can also be recognized as "outlier" image elimination. Here, image consistency detection can complement occlusion detection. When occlusion detection fails, image consistency detection can also remove images containing foreground objects from the visible image list. As shown in Figure 7, the architectural texture pasting errors of the texture-mapping results of the dataset TJH-008812 are compared, and the pavement part containing architectural error texture is enlarged and displayed. It should be the case that for some areas of the lower right building, the occlusion detection algorithm fails, resulting in the wrong building texture on the road surface. It can be seen that compared with the texture-mapping results using the image consistency detection algorithm based on the mean shift, this paper uses the texture-mapping method based on BRIEF image consistency detection to avoid the building paste error when occlusion detection fails. The semantic accuracy of the road area texture-mapping results is high, which shows the effectiveness of this algorithm.

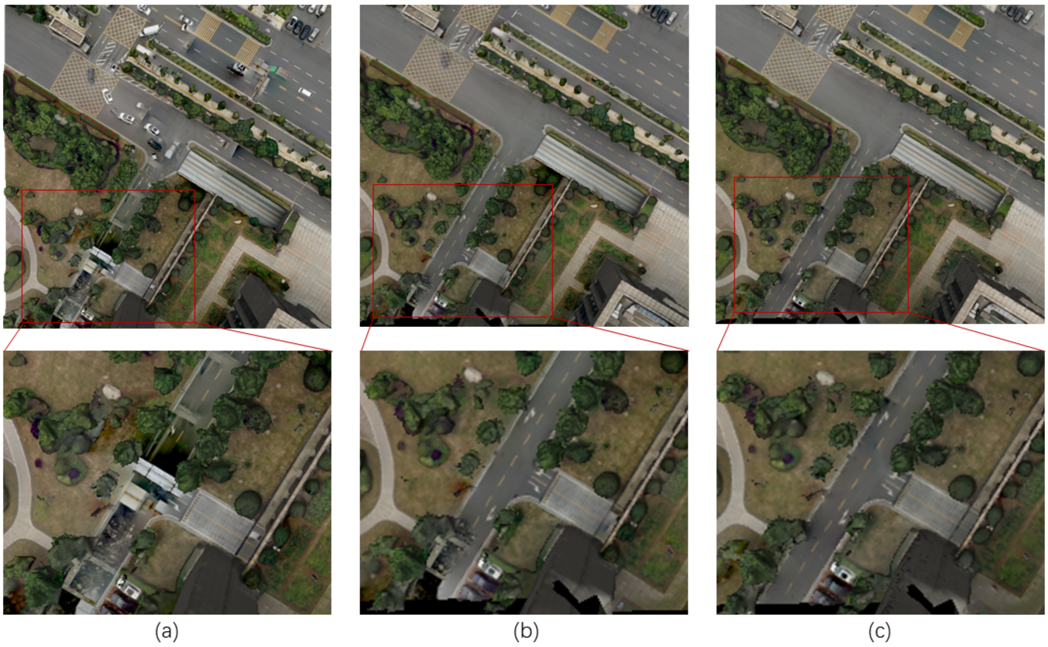

Figure 7. Results comparison without occlusion detection model. (**a**) results of experiments without image consistency detection; (**b**) results of the method of mvs-texturing; and (**c**) results of our method.

Figure 8 compares the texture-mapping results for the three methods on the dataset TJH-0078. Both methods in this article were better than mvs-texturing in terms of removing moving vehicle error textures, and the view selection method with the vehicle semantic information worked best. The limitation of the proposed method with semantic information was the accuracy of the vehicle mask image. This method avoided almost all moving vehicle error textures as long as the acquired vehicle mask image was sufficiently accurate. In this article, the accuracy of the manually segmented vehicle mask images was high, and as a result, the effect of removing moving vehicle error textures was better. The BRIEF-based image consistency detection method is based on the color information within the mesh triangle projection to eliminate the "outlier" images. It was not effective in removing some of the moving vehicles whose color values were close to the road surface. Therefore, the BRIEF-based image consistency detection method was not as good as the method with semantic information of vehicles in reducing moving vehicle error textures on urban roads. The proposed BRIEF-based image consistency detection method, however, effectively reduced the error of objects not reconstructed, such as streetlights and powerlines. It also reduced the error textures of objects whose detection failed because occlusion occurred at the same time and could improve the texture-mapping quality in many aspects. The view selection method with semantic information was ineffective in reducing other erroneous textures, such as power lines and telegraph poles, because it considered only specific semantic information, just as other deep learning methods [31,32].

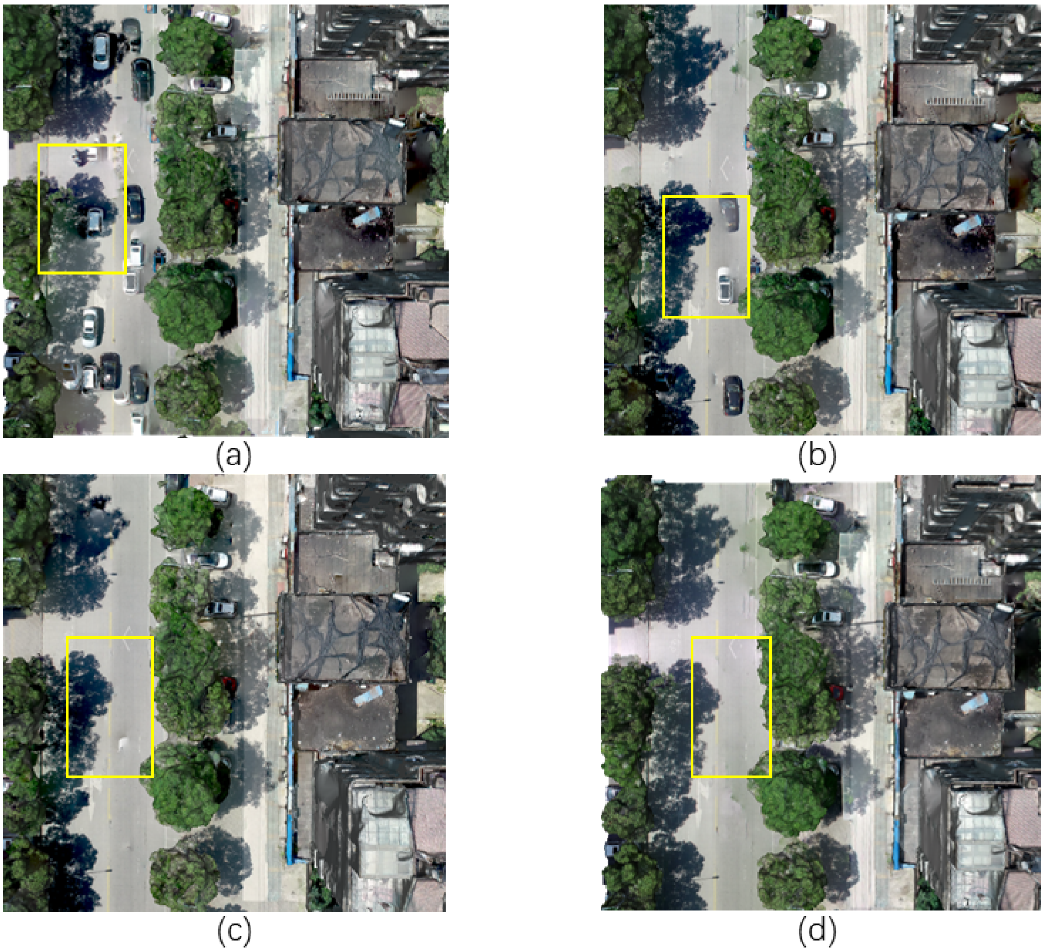

Figure 8. Comparison of texture-mapping results of the road wires of in dataset TJH-0078. (**a**) results of experiments without image consistency detection; (**b**) results of mvs texturing; (**c**) results of our method; (**d**) results of our method with semantic information.

4. Conclusions

In image-based 3D reconstruction, unreconstructed objects can lead to the problem of inconsistency between model geometry and texture, resulting in texture-mapping errors. To address this problem, firstly, we uniformly sample the color values of the triangular surface and its adjacent triangular surface within the projection range of the visible image, and we calculate a BRIEF descriptor for each image according to the sampled color values. Then, the Hamming distance between the corresponding descriptors of the visible image is calculated to remove the "outlier" images. The experiments showed that the BRIEF descriptors had stronger feature description ability than the color averages used in existing works. Thus, the outlier detection method better removed the noisy images in the image list and reduced the probability of the noisy images being selected in the view selection stage. At the same time, adding the semantic information to the MRF energy function also effectively reduced the probability of the noisy images being selected. Thus, the errors of inconsistent geometric structure and texture information caused by unreconstructed objects can be removed effectively, which is important for improving the texture quality of realistic 3D models. However, when faced with some small objects, if the sampling point does not fall on the small objects, or the object detection algorithm does not detect it, then the algorithm in this paper may fail; this is also the shortcoming of the algorithm in this paper, which we will improve in the future.

Author Contributions: Conceptualization, J.Y. and L.L.; methodology, J.Y.; software, L.L.; validation, L.L. and G.P.; formal analysis, L.L. and G.P.; investigation, H.H.; writing—original draft preparation, L.L.; writing—review and editing, J.Y.; visualization, L.L.; supervision, L.L., and G.P.; project administration, F.D.; funding acquisition, J.W. and J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, Grant Numbers 42201483 and 41874029, and China Postdoctoral Science Foundation, Grant Number 2022M710332.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank Pei Tao, Yingjie Qu, and Qingsong Yan for their valuable suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Snavely, N.; Seitz, S.M.; Szeliski, R. Photo tourism: Exploring photo collections in 3D. In Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference, Boston, MA, USA, 30 July–3 August 2006; pp. 835–846.
2. Schonberger, J.L.; Frahm, J.M. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113.
3. Wu, C. Towards linear-time incremental structure from motion. In Proceedings of the 2013 International Conference on 3D Vision-3DV 2013, Seattle, WA, USA, 29 June 2013–1 July 2013; pp. 127–134.
4. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *32*, 1362–1376. [CrossRef] [PubMed]
5. Shen, S. Accurate multiple view 3d reconstruction using patch-based stereo for large-scale scenes. *IEEE Trans. Image Process.* **2013**, *22*, 1901–1914. [CrossRef] [PubMed]
6. Yao, Y.; Luo, Z.; Li, S.; Zhang, J.; Ren, Y.; Zhou, L.; Fang, T.; Quan, L. Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1790–1799.
7. Labatut, P.; Pons, J.P.; Keriven, R. Robust and efficient surface reconstruction from range data. In *Computer Graphics Forum*; Blackwell Publishing Ltd.: Oxford, UK, 2009; Volume 28, pp. 2275–2290.
8. Kazhdan, M.; Hoppe, H. Screened poisson surface reconstruction. *ACM Trans. Graph. (ToG)* **2013**, *32*, 1–13. [CrossRef]
9. Rothermel, M.; Gong, K.; Fritsch, D.; Schindler, K.; Haala, N. Photometric multi-view mesh refinement for high-resolution satellite images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 52–62. [CrossRef]
10. Waechter, M.; Moehrle, N.; Goesele, M. Let there be color! Large-scale texturing of 3D reconstructions. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 836–850.
11. Bernardini, F.; Martin, I.M.; Rushmeier, H. High-quality texture reconstruction from multiple scans. *IEEE Trans. Vis. Comput. Graph.* **2001**, *7*, 318–332. [CrossRef]
12. Callieri, M.; Cignoni, P.; Corsini, M.; Scopigno, R. Masked photo blending: Mapping dense photographic data set on high-resolution sampled 3D models. *Comput. Graph.* **2008**, *32*, 464–473. [CrossRef]
13. Allene, C.; Pons, J.P.; Keriven, R. Seamless image-based texture atlases using multi-band blending. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008; pp. 1–4.
14. Lempitsky, V.; Ivanov, D. Seamless mosaicing of image-based texture maps. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–6.
15. Yang, J.; Liu, L.; Xu, J.; Wang, Y.; Deng, F. Efficient global color correction for large-scale multiple-view images in three-dimensional reconstruction. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 209–220. [CrossRef]
16. Li, S.; Xiao, X.; Guo, B.; Zhang, L. A Novel OpenMVS-Based Texture Reconstruction Method Based on the Fully Automatic Plane Segmentation for 3D Mesh Models. *Remote Sens.* **2020**, *12*, 3908. [CrossRef]
17. Fu, Y.; Yan, Q.; Liao, J.; Xiao, C. Joint texture and geometry optimization for RGB-D reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5950–5959.
18. Bi, S.; Kalantari, N.K.; Ramamoorthi, R. Patch-based optimization for image-based texture mapping. *ACM Trans. Graph.* **2017**, *36*, 106–111. [CrossRef]
19. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106. [CrossRef]

20. Barron, J.T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P.P. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 5855–5864.
21. Müller, T.; Evans, A.; Schied, C.; Keller, A. Instant neural graphics primitives with a multiresolution hash encoding. *arXiv* **2022**, arXiv:2201.05989.
22. Martin-Brualla, R.; Radwan, N.; Sajjadi, M.S.; Barron, J.T.; Dosovitskiy, A.; Duckworth, D. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Montreal, QC, Canada, 11–17 October 2021; pp. 7210–7219.
23. Rudnev, V.; Elgharib, M.; Smith, W.; Liu, L.; Golyanik, V.; Theobalt, C. Nerf for outdoor scene relighting. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 615–631.
24. Zhang, X.; Srinivasan, P.P.; Deng, B.; Debevec, P.; Freeman, W.T.; Barron, J.T. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Trans. Graph. (TOG)* **2021**, *40*, 1–18. [CrossRef]
25. Srinivasan, P.P.; Deng, B.; Zhang, X.; Tancik, M.; Mildenhall, B.; Barron, J.T. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 7495–7504.
26. Sinha, S.N.; Steedly, D.; Szeliski, R.; Agrawala, M.; Pollefeys, M. Interactive 3D architectural modeling from unordered photo collections. *ACM Trans. Graph. (TOG)* **2008**, *27*, 1–10. [CrossRef]
27. Grammatikopoulos, L.; Kalisperakis, I.; Karras, G.; Petsa, E. Automatic multi-view texture mapping of 3D surface projections. In Proceedings of the 2nd ISPRS International Workshop 3D-ARCH, Zurich, Switzerland, 12–13 July 2007, pp. 1–6.
28. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
29. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
30. Zhu, Q.; Shang, Q.; Hu, H.; Yu, H.; Zhong, R. Structure-aware completion of photogrammetric meshes in urban road environment. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 56–70. [CrossRef]
31. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Volume 28.
32. Yang, C.; Zhang, F.; Gao, Y.; Mao, Z.; Li, L.; Huang, X. Moving Car Recognition and Removal for 3D Urban Modelling Using Oblique Images. *Remote Sensing* **2021**, *13*, 3458. [CrossRef]
33. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. Brief: Binary robust independent elementary features. In Proceedings of the European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; pp. 778–792.
34. Boykov, Y.; Veksler, O.; Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1222–1239. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.